

PATENT  
ATTY. DOCKET NO. IBM/265RECEIVED  
CENTRAL FAX CENTER

Applicant : Abdo Esmail Abdo et al. Art Unit: 2162  
Serial No. : 10/758,486 Examiner: Giovanna B. Colan  
Filed : January 15, 2004  
For : GENERATING STATISTICS ON TEXT PATTERN MATCHING  
PREDICATES FOR ACCESS PLANNING

JAN 14 2008

Mail Stop Amendment  
Commissioner of Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Via Facsimile

RESPONSE TO NOTICE OF NON-COMPLIANT APPEAL BRIEF

This paper and the attached Appeal Brief is in response to the Notice of Non-Compliant Appeal Brief mailed December 13, 2007.

If any petition for extension of time is necessary to accompany this communication, please consider this paper a petition for such an extension of time, and apply the appropriate extension of time fee to Deposit Account 23-3000. If any other charges or credits are necessary to complete this communication, please apply them to Deposit Account 23-3000.

Respectfully submitted,




Thomas W. Humphrey  
Reg. No. 34,353

Wood, Herron & Evans, L.L.P.  
2700 Carew Tower  
441 Vine Street  
Cincinnati, OH 45202-2917  
Voice: (513) 241-2324/Facsimile: (513) 241-6234

Certificate of Mailing or Facsimile

I hereby certify that this correspondence and any enclosures are being deposited with the United States Postal Service via U.S. Mail, Prepaid First Class Postage or transmitted via facsimile to the number as provided by the United States Patent and Trademark Office on the date indicated below.

  
Thomas W. Humphrey  
Reg. No. 34,353January 14, 2008  
Date

**RECEIVED  
CENTRAL FAX CENTER****JAN 14 2008****PATENT  
ATTY. DOCKET NO. IBM/265**

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE  
BEFORE THE BOARD OF PATENT APPEALS  
AND INTERFERENCES**

**Ex parte Abdo Esmail Abdo et al.****Appeal No. \_\_\_\_\_**

**Serial No.: 10/758,486  
Filed: January 15, 2004  
Group Art Unit: 2162  
Examiner: Giovanna B. Colan  
Applicant: Abdo Esmail Abdo et al.  
Title: GENERATING STATISTICS ON TEXT PATTERN  
MATCHING PREDICATES FOR ACCESS PLANNING**

Cincinnati, Ohio 45202

Original Submission September 24, 2007  
Resubmitted January 14, 2008  
*Via Facsimile*

**APPEAL BRIEF**

This brief is in furtherance of Applicant's Notice of Appeal filed July 23, 2007,  
appealing the decision of the Examiner dated February 23, 2007 finally rejecting claims 1-23.  
A copy of the claims appears in the Appendix to this brief.

---

**Certificate of Facsimile or Mail Transmission**

I hereby certify that this correspondence and any enclosures are being deposited with the United States Postal Service via U.S. Mail, Prepaid First Class Postage or transmitted via facsimile to the number as provided by the United States Patent and Trademark Office on the date indicated below.

/Thomas W. Humphrey/ January 14, 2008  
Thomas W. Humphrey Date  
Reg. No. 34,353

**Real Party In Interest**

The real party in interest in this appeal is INTERNATIONAL BUSINESS MACHINES CORPORATION, a corporation of New York having a place of business at New Orchard Road, Armonk, New York 10504.

**Related Appeals and Interferences**

There are no such appeals or interferences.

**RECEIVED  
CENTRAL FAX CENTER****JAN 14 2008****Status of Claims****Total Number of Claims in the Application**

Claims in the application are: 23

**Status of all the Claims**

1. Claims canceled: NONE
2. Claims withdrawn from consideration but not canceled: NONE
3. Claims objected to: NONE
4. Claims allowed or confirmed: NONE
5. Claims rejected: 1-23

**Claims on Appeal**

The claims are appeal are Claims 1-23. The claims in this application have never been amended.

**Status of Amendments**

An Amendment of Appeal filed with this brief, correcting typographical errors and omissions in the specification, is pending.

**Summary of Claimed Subject Matter as to Independent Claim 1**

The subject matter of independent Claim 1 is described in the specification on page 9, line 19 to page 12, line 11, and Figs. 4-5, reference numbers 200-350 of the drawings.

These pages disclose examples of the invention recited in claim 1, which describes a method that utilizes “a data storage structure storing character statistics on an attribute” (see Fig. 4 and the description thereof at page 9 lines 22-25 and page 10 lines 1-9), comprising “a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position” (see in Fig. 4, the storage of frequently occurring characters for positions 1, 2, 3 and 4). This structure is used in conjunction with a “pattern matching predicate” (e.g., as noted at page 12, line 12 to page 14 line 20, which explains the process in conjunction with Fig. 5, and the example at page 14 line 21 - page 16 line 6 relating to the pattern match to the string “BIB”). As recited in the claim, pattern matching with the structure proceeds by the use of “character positions of characters in [the] pattern matching predicate” as can be seen in the process of Fig. 6 and the example noted above.

Thus, the invention is directed to the storage of statistical information regarding characters in an attribute of a database relation, and the use of those statistics in evaluating a pattern matching predicate.

**Summary of Claimed Subject Matter as to Independent Claim 11**

Independent Claim 11 is directed to a computer system, such as that described in the specification at page 5 line 4 through page 9 line 18, which implements a process as described on page 9, line 19 to page 12, line 11, and Figs. 4-5, reference numbers 200-350 of the drawings.

The latter pages disclose examples of the invention recited in claim 11, which describes an apparatus that utilizes “a data storage structure storing character statistics on an attribute” (see Fig. 4 and the description thereof at page 9 lines 22-25 and page 10 lines 1-9), comprising “a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position” (see in Fig. 4, the storage of frequently occurring characters for positions 1, 2, 3 and 4). This structure is used in conjunction with a “pattern matching predicate” (e.g., as noted at page 12, line 12 to page 14 line 20, which explains the process in conjunction with Fig. 5, and the example at page 14 line 21 - page 16 line 6 relating to the pattern match to the string “BIB”). As recited in the claim, pattern matching with the structure proceeds by the use of “character positions of characters in [the] pattern matching predicate” as can be seen in the process of Fig. 6 and the example noted above.

Thus, the invention of claim 11 is also directed to the storage of statistical information regarding characters in an attribute of a database relation, and the use of those statistics in evaluating a pattern matching predicate.



**RECEIVED  
CENTRAL FAX CENTER****JAN 14 2008****Summary of Claimed Subject Matter as to Independent Claim 21**

Independent Claim 21 is directed to a program product, implementing a relational database system such as described in the specification on page 9, line 19 to page 12, line 11, and Figs. 4-5, reference numbers 200-350 of the drawings.

These pages disclose examples of the invention recited in claim 21, which describes a program product that utilizes "a data storage structure storing character statistics on an attribute" (see Fig. 4 and the description thereof at page 9 lines 22-25 and page 10 lines 1-9), comprising "a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position" (see in Fig. 4, the storage of frequently occurring characters for positions 1, 2, 3 and 4). This structure is used in conjunction with a "pattern matching predicate" (e.g., as noted at page 12, line 12 to page 14 line 20, which explains the process in conjunction with Fig. 5, and the example at page 14 line 21 - page 16 line 6 relating to the pattern match to the string "BIB"). As recited in the claim, pattern matching with the structure proceeds by the use of "character positions of characters in [the] pattern matching predicate" as can be seen in the process of Fig. 6 and the example noted above.

Thus, again, claim 21 is directed to the storage of statistical information regarding characters in an attribute of a database relation, and the use of those statistics in evaluating a pattern matching predicate.

**Grounds of Rejection under 36 U.S.C. 103(a)**

Claims 1-23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Fritchman (US Patent No. 6,785,677 B1, filed: May 2, 2001) in view of Haas et al. (Haas hereinafter) (Non-Patent Literature: Sampling – Based Selectivity for Joins Using Augmented Frequent Value Statistics, IBM Almaden Research Center, San Jose CA, 1995, IEEE).

Argument

The Examiner's rejection is premised upon the assertion that the Fritchman '677 patent discloses storing character structures that meet the claim language. The claim language recites a structure in which, "for each of a plurality of character positions" the structure identifies "frequently occurring characters in that character position, and statistics for each frequently occurring character". Fritchman does not disclose this. Fritchman, in the cited passages, only discloses how to describe a character string criterion for a query, indicating that strings should be described using a prefix segment, a suffix segment, and one or more interior segments.

The Examiner has cited to passages in col. 8 of Fritchman regarding a "bit vector noting the position of any single-character wild cards" (that can be included in a LIKE predicate of a query). The Examiner seems to believe that the representation of a LIKE predicate using wildcard characters, is somehow the same as a statistical representation of frequently occurring characters in particular character locations in an attribute. They are clearly not the same. A wildcard is a way to identify a pattern of characters that match a criterion, not a way of identifying, for an attribute of a relation, frequently occurring characters in a character position. Indeed, Fritchman does not disclose the generation of any such statistical information, or the use of the same, but rather discloses a matching methodology in which a "pattern is applied to each database value in turn, where the value is the result of evaluating the first operand of the LIKE predicate for each candidate row retrieved from the database." Clearly, the application of a pattern to every database value to

determine exactly how many matches exist, is entirely different from using characterizing data to estimate a number of matches.

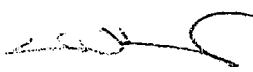
The Examiner asserts the Haas paper discloses storing character statistics on an attribute, and statistics for each frequently occurring character. However, the Examiner is incorrect here as well. The Examiner asserts that page 528 of Haas discloses “storing character statistics on an attribute” and “statistics for each frequently occurring character”. Applicant is unable to identify any such disclosure. Specifically, the text quoted in Haas is directed to sampling-based selectivity estimation, using what are called ‘augmented frequent value’ (AFV) statistics. While selectivity estimation and the use of frequent value statistics is somewhat relevant to the present invention, Applicant has been unable to identify any disclosure in the cited paragraphs of the creation of an index or statistic for text matching, that stores “for each of a plurality of character positions, frequently occurring characters in that character position”, as claimed. The quoted paragraphs on page 528 only generally discuss the creation of AFV statistics and do not detail a particular form of such statistics for text matching, much less the particular form that is recited in the present claims, namely, “for each of a plurality of character positions, frequently occurring characters in that character position.”

As the Examiner has not identified how either cited reference relied upon in the rejection discloses a index or statistic meeting the claim recitations, namely, “for each of a plurality of character positions, frequently occurring characters in that character position,” Applicant submits that all claims presented herein are allowable and requests the issuance of a Notice of Allowability.

The Examiner attention is directed to an Amendment for Appeal attached hereto which corrects typographical errors noted by Applicant during the preparation of this appeal brief. Applicant submits that these amendments do not introduce new matter or new issues and respectfully requests their entry.

In accordance with the foregoing, Applicant submits that the Examiner's rejection is in error and a reversal of the rejection and allowance of the claims is therefore requested.

Respectfully submitted,  
Wood, Herron & Evans, L.L.P.

By:   
Thomas W. Humphrey  
Reg. No. 34,353

2700 Carew Tower  
441 Vine Street  
Cincinnati, OH 45202-2917

Voice: (513) 241-2324  
Facsimile: (513) 241-6234

**RECEIVED  
CENTRAL FAX CENTER****JAN 14 2008****Claim Appendix**

1. (original) A method for generating a statistic for a pattern matching predicate on an attribute of a relation, to be used in optimizing execution of a query directed to one or more attributes of said relation, comprising  
providing a data storage structure storing character statistics on said attribute, comprising a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position, and statistics for each frequently occurring character,  
retrieving said statistics in response to said pattern matching predicate based upon the character positions of characters in said pattern matching predicate, and  
generating said statistic based upon said retrieved character statistics.
2. (original) The method of claim 1 wherein said data storage structure further stores a count of a number of occurrences of a frequently occurring character.
3. (original) The method of claim 1 wherein said data storage structure further comprises a second structure storing frequently occurring characters that are subsequent to the frequently occurring characters stored in said first structure.
4. (original) The method of claim 3 wherein said second data storage structure further stores a probability of occurrence of a frequently occurring subsequent character.

5. (original) The method of claim 1 wherein generating said statistic based upon said retrieved character statistics comprises estimating a statistic for a desired character and desired character position, in the event said statistic is not stored in said first structure.

6. (original) The method of claim 5 wherein said estimating comprises accumulating statistics for said desired character position, for characters other than said desired character, and calculating average frequency of occurrence of characters that do not have statistics stored in said first structure.

7. (original) The method of claim 1 wherein said first structure is a table.

8. (original) The method of claim 1 wherein said first structure comprises a linked list.

9. (original) The method of claim 3 wherein said second structure comprises a table.

10. (original) The method of claim 3 wherein said second structure comprises a linked list.

11. (original) A computer system implementing a relational database system and generating a statistic for a pattern matching predicate on an attribute of a

relation of said relational database, to be used in optimizing execution of a query directed to one or more attributes of said relation, comprising

storage for said relational database, including a relation having a plurality of tuples including values for a plurality of attributes, and a data storage structure storing character statistics on an attribute, comprising a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position, and statistics for each frequently occurring character, and

computing circuitry performing query optimization and query execution upon said relational database, said query optimization including generating said statistic for an attribute of said relation by retrieving said statistics in response to said pattern matching predicate based upon the character positions of characters in said pattern matching predicate, and generating said statistic based upon said retrieved character statistics.

12. (original) The computer system of claim 11 wherein said data storage structure further stores a count of a number of occurrences of a frequently occurring character.

13. (original) The computer system of claim 11 wherein said data storage structure further comprises a second structure storing frequently occurring characters that are subsequent to the frequently occurring characters stored in said first structure.



14. (original) The computer system of claim 13 wherein said second data storage structure further stores a probability of occurrence of a frequently occurring subsequent character.

15. (original) The computer system of claim 11 wherein said computing circuitry generates said statistic based upon said retrieved character statistics by estimating a statistic for a desired character and desired character position, in the event said statistic is not stored in said first structure.

16. (original) The computer system of claim 15 wherein said computer circuitry performs said estimating by accumulating statistics for said desired character position, for characters other than said desired character, and calculating average frequency of occurrence of characters that do not have statistics stored in said first structure.

17. (original) The computer system of claim 11 wherein said first structure is a table.

18. (original) The computer system of claim 11 wherein said first structure comprises a linked list.

19. (original) The computer system of claim 13 wherein said second structure comprises a table.

20. (original) The computer system of claim 13 wherein said second structure comprises a linked list.

21. (original) A program product for implementing a relational database system and generating a statistic for a pattern matching predicate on an attribute of a relation of said relational database, to be used in optimizing execution of a query directed to one or more attributes of said relation, comprising

a relational database, including a relation having a plurality of tuples including values for a plurality of attributes,

a data storage structure storing character statistics on an attribute, comprising a first structure storing, for each of a plurality of character positions, frequently occurring characters in that character position, and statistics for each frequently occurring character, and

relational database software performing query optimization and query execution upon said relational database, said query optimization including generating a statistic for an attribute of said relation by retrieving said statistics in response to said pattern matching predicate based upon the character positions of characters in said pattern matching predicate, and generating said statistic based upon said retrieved character statistics, and

a signal bearing media holding said relational database and relational database software.

22. (original) The program product of claim 21 wherein the signal bearing media comprises transmission media.

23. (original) The program product of claim 21 wherein the signal bearing media comprises recordable media.

**Evidence Appendix**

None.

**Related Proceedings Appendix**

None.

## TABLE OF CONTENTS

Real Party In Interest .....	-2-
Related Appeals and Interferences .....	-3-
Status of Claims .....	-4-
Total Number of Claims in the Application .....	-4-
Status of all the Claims .....	-4-
Claims on Appeal .....	-4-
Status of Amendments .....	-5-
Summary of Claimed Subject Matter as to Independent Claim 1 .....	-6-
Summary of Claimed Subject Matter as to Independent Claim 11 .....	-7-
Summary of Claimed Subject Matter as to Independent Claim 21 .....	-8-
Grounds of Rejection under 36 U.S.C. 103(a) .....	-9-
Argument .....	-10-
Claim Appendix .....	-13-
Evidence Appendix .....	-19-
Related Proceedings Appendix .....	-20-